

# Ein Revenue-Management-System zur multikriteriellen Kapazitätssteuerung im Umfeld des Cloud Computing

Michael Mohaupt  
Andreas Hilbert

Veröffentlicht in:  
Multikonferenz Wirtschaftsinformatik 2012  
Tagungsband der MKWI 2012  
Hrsg.: Dirk Christian Mattfeld; Susanne Robra-Bissantz



Braunschweig: Institut für Wirtschaftsinformatik, 2012

# Ein Revenue-Management-System zur multikriteriellen Kapazitätssteuerung im Umfeld des Cloud Computing

**Michael Mohaupt, Andreas Hilbert**

Technische Universität Dresden, Professur für Wirtschaftsinformatik,  
insbes. Informationssysteme im Dienstleistungsbereich,  
E-Mail: Michael.Mohaupt@mailbox.tu-dresden.de bzw. Andreas.Hilbert@tu-dresden.de

## Abstract

Die effiziente Nutzung beschränkter Kapazitätsressourcen ist ein kritischer Erfolgsfaktor für Cloud-Anbieter. Revenue-Management-Systeme steuern daher die eintreffenden Kundenanfragen unterschiedlicher Wertigkeit, um die Erlöse zu maximieren. Während die Annahme der Anfrage einen positiven Einfluss auf die Kundenloyalität haben kann, ist bei der Ablehnung deren Verringerung zu erwarten. Das traditionelle Optimierungsmodell, das nur eine Maximierung der kurzfristigen Transaktionserlöse vorsieht, wird daher um loyalitätsbeeinflussende Steuerungsentscheidungen erweitert, die sich auch auf die zukünftigen Umsätze und Cross Selling (d.h. langfristige Erfolgspotenzial) auswirken können. Eine Simulation des erweiterten Informationssystems in MATLAB wird zur Effizienzbeurteilung durchgeführt.

## 1 Problemstellung

Viele von Unternehmen genutzte IT-Anwendungen (z.B. Data Mining, Simulationen) und Dienste (z.B. Hosting) unterliegen einem stark schwankenden Bedarf an IT-Ressourcen. Zur Sicherstellung ihrer Verfügbarkeit sehen sich Unternehmen hohen Investitionen in eine adäquate IT-Infrastruktur gegenüber [13]. Der steigende Wettbewerbsdruck erfordert aber zudem die Konzentration auf die Kernkompetenzen zur Entwicklung neuer Produkte und Leistungen für Kunden und der Optimierung eigener Prozesse – unter der Herausforderung, die IT-Kosten so gering wie möglich zu halten [4]. Cloud Computing als vielversprechender Ansatz hat sowohl von Forschern als auch Praktikern ein steigendes Interesse erfahren, da es die Deckung dieses stark schwankenden Bedarfs der IT-Nutzer nach IT-Ressourcen zu angemessenen Preisen ermöglicht [2]. Cloud Computing soll hier als ein Pool aus einfach benutzbaren und zugänglichen virtualisierten Ressourcen (wie Hardware, Entwicklungsplattformen, Diensten) verstanden werden, die jeweils dynamisch an den schwankenden Bedarf angepasst werden können [16]. In sogenannten Clouds werden IT-Ressourcen wie CPU, Arbeitsspeicher, Speicherplatz und Bandbreite in einzelne Dienste gebündelt, die dann den Cloud-Benutzern zu unterschiedlichen Preisen angeboten werden. Durch die Inanspruchnahme jener Dienste können beim Cloud-Kunden zwischen 60 bis 80% der IT-Kosten eingespart werden [8]. Cloud-Anbieter sehen sich im Marktumfeld

einer unsicheren, zeitlich verteilten Nachfrage nach Dienstklassen unterschiedlicher Wertigkeit gegenüber [2]. Dies bedingt eine nicht-triviale Entscheidung über die Annahme bzw. Ablehnung von Buchungsanfragen. Auch weil das Risiko einer Fehleinschätzung des (zukünftigen) IT-Bedarfs vom Kunden auf den Anbieter übertragen wird [4], ist für letztere die effiziente Ressourcennutzung der beschränkten Ressourcen ein kritischer Erfolgsfaktor [17]. Im Rahmen des Revenue Managements (Kapazitätssteuerung) wird daher versucht, der Zielstellung einer erlösoptimalen Ressourcenallokation mit komplexen Steuerungsmethoden des Operations Research unter Mithilfe von informationstechnologischen Anwendungssystemen und unter Berücksichtigung einer breiten Datenbasis zu begegnen [15]. Da insbesondere die Kunden im Cloud Computing der ständigen Verfügbarkeit der IT-Ressourcen eine sehr hohe Bedeutung beimessen [16], können sich die Steuerungsentscheidungen des Anbieters zur Verfügbarkeit des angefragten Dienstes als Folge auf die Loyalität der Nachfrager auswirken. Während die Annahme einer Buchungsanfrage einen positiven Einfluss auf die Kundenloyalität haben kann, ist durch die Ablehnung der Dienstanfrage eine Gefährdung der Loyalität nicht ausgeschlossen [11], [18]. Diese Loyalitätsveränderungen werden sich dann auf den zukünftigen Umfang der Transaktionen und des Cross Selling niederschlagen [6]. Um langfristig Gewinne zu maximieren, sollten Cloud-Anbieter daher loyale Beziehungen mit Potenzial- (geringe aktuelle aber hohe zukünftige Wertbeiträge) und Referenzkunden (geringe direkte aber hohe indirekte Wertbeiträge) aufbauen [17]. Allerdings werden im Rahmen des traditionellen Optimierungsmodells nur die transaktionsorientierten Erlöse basierend auf der Zahlungsbereitschaft der Kunden maximiert. Die Steuerungsentscheidungen des Anbieters, die die Loyalität und somit zukünftige Erlöse beeinflussen können, werden nicht berücksichtigt. Als Folge kann der Aufbau von Beziehungen mit langfristig profitablen Kunden gefährdet sein. Nach den Ausführungen zum Revenue-Management-System und zur Kundenloyalität soll daher ein erweitertes Optimierungsmodell entwickelt und dieses im Rahmen einer Simulation mit dem traditionellen Modell verglichen werden.

## 2 Revenue-Management-System im Cloud Computing

Für den Cloud-Anbieter ergibt sich durch die Beteiligung des Kunden eine unsichere Einflussgröße bezüglich Umfang, Wert und zeitlichem Eintreffen der Nachfrage sowie der Kundenreaktion, falls das angefragte Angebot nicht verfügbar ist [15]. Des Weiteren bewirkt diese Integration, dass die angefragten Cloud-Dienste im Moment des Leistungskonsums erstellt werden und damit nicht lagerfähig sind. Folglich kann der Verzicht auf eine geeignete Kapazitätssteuerung dazu führen, dass ein Großteil der Kapazität für frühe, niederwertige Buchungsanfragen reserviert, die verfügbare Kapazität überbucht wird oder nicht in Anspruch genommene Reservierungen Leerkosten verursachen [17].

Die Bündelung der Kapazitätsressourcen zu Diensten mit unterschiedlicher Wertigkeit führt dazu, dass die Dienste um Ressourcen konkurrieren. Der Cloud-Anbieter muss daher sorgfältig über die Verwendung der beschränkten Kapazität entscheiden [13]. Trotz fehlender Kenntnis über den Umfang und den Wert der zukünftigen Anfragen muss bei Eintreffen einer Buchungsanfrage über deren Annahme bzw. Ablehnung entschieden werden. Bei Annahme einer niederwertigen Anfrage besteht die Gefahr, dass eine später eintreffende, höherwertige Anfrage mangels verfügbarer Kapazität abgelehnt werden muss (sog. Umsatzverdrängung). Im Fall der Ablehnung der niederwertigen Anfrage kann ein Umsatzverlust resultieren, falls solche höherwertigen Anfragen ausbleiben [15]. Die geeignete Steuerung, d.h. dem Abwägen zwischen

diesen entgegengesetzten erlösrelevanten Auswirkungen, stellt dabei den Gegenstand des Revenue Managements dar. Das Ziel der effizienten Nutzung der operativ unflexiblen Kapazität wird dabei über die (Gesamt)-Erlösmaximierung operationalisiert.

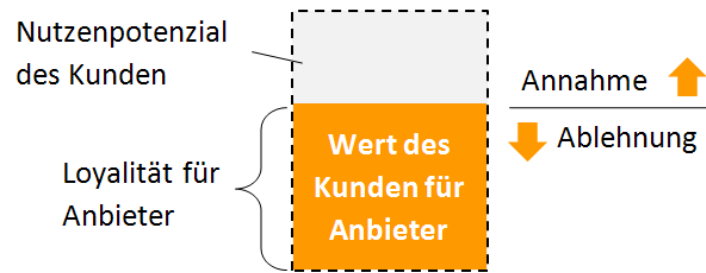
Die beschriebene Problematik im Revenue Management ist zwar grundsätzlich einer formal-quantitativen Modellierung und (teil)-automatischen Lösung zugänglich, erweist sich jedoch als sehr komplex, da die Merkmale optimierungswürdiger Entscheidungsprobleme (d.h. hinreichend kompliziert, datenintensiv und schnell zu entscheiden) erfüllt sind [9]. So müssen Revenue-Management-Systeme bspw. im Millisekundenbereich über die Annahme von Anfragen entscheiden und die Optimierungsaufgaben unterliegen oftmals solch hoher Berechnungsintensität, dass diese meist nur vor oder zu ausgewählten Zeitpunkten in der Buchungsperiode durchgeführt werden können [15]. Revenue-Management-Systeme können damit als intelligente Informationssysteme begriffen werden [12], die den Anwender bei der erlösoptimalen, (teil)-automatischen Ressourcenallokation entscheidend unterstützen. Jene Informationssysteme, die in der Lage sind, die richtigen Informationen zu sammeln und an Entscheidungspunkten zeitnah zur Verfügung zu stellen, können sich als kritischer Erfolgsfaktor erweisen und Vorteile gegenüber Wettbewerbern verschaffen [17].

Sollen auch die Auswirkungen der Steuerungsentscheidungen im Rahmen des Revenue Managements Berücksichtigung finden, so müssen auf strategischer Ebene neben den kapazitäts- bzw. transaktionsorientierten auch kundenbeziehungsorientierte Ziele formuliert werden. So kann zusätzlich zum (kurzfristigen) Erfolg durch die Maximierung der Erlöse aus der Kapazitätsnutzung auch das (langfristige) Erfolgspotenzial durch die Entwicklung profitabler Kundenbeziehungen etabliert werden. Im nächsten Abschnitt soll daher genauer auf die (langfristige) Kundenwertigkeit und -loyalität eingegangen werden.

### 3 Kundenloyalität und Kundenwert

Der Wert eines Cloud-Kunden für den Anbieter ist nicht auf den Gewinn der einzelnen Transaktion beschränkt, sondern ergibt sich als der gesamte Gewinn, den der Kunde während der Geschäftsbeziehung mit dem Anbieter beisteuert. Der Kundenwert (als Wert des Kunden aus Unternehmenssicht) stellt dem beziehungsorientierten Marketingparadigma zufolge eine zentrale Einflussgröße auf Managemententscheidungen dar und ist eng mit dem Shareholder Value verbunden [17]. Die Vielschichtigkeit der Wertbeiträge der Kunden und die daraus resultierende Variation der Kundenwerte und deren Einflussfaktoren erfordern daher eine entsprechende Differenzierung der Unternehmensaktivitäten. Der Kundenwert soll hier als das Produkt zweier Bestandteile verstanden werden (vgl. Bild 1, S. 4). Zunächst das vom Anbieter unabhängige Nutzenpotenzial des Kunden als Resultat der möglichen Inanspruchnahme der Leistungen *aller* Anbieter in zukünftigen Buchungsperioden. So beabsichtigt bspw. ein Cloud-Kunde 5000 € für IT-Dienste in den nächsten 12 Monaten auszugeben. Den zweiten Faktor bildet die gegenwärtige Loyalität des Kunden, d.h. die Wahrscheinlichkeit, dass die Beziehung mit dem Anbieter fortbestehen wird [6]. Dieser Bestandteil repräsentiert damit den Anteil des Nutzenpotenzials, den der Anbieter durch seine Leistungsbereitschaft beim Kunden abschöpfen kann. Den möglichen verbleibenden Rest teilt sich der Wettbewerb. Anbieter ohne (Loyalitäts-)Bindung gehen leer aus. Kundenloyalität äußert sich somit in einem grundsätzlichen Vertrauensverhältnis und einer allgemein positiven Einstellung des Kunden gegenüber dem Unternehmen. Die Kundenzufriedenheit (als Antezedenz der Loyalität) übt hier einen positiven Einfluss aus [7]. Wann immer

also die Steuerungsentscheidungen des Anbieters zur Verfügbarkeit der angefragten IT-Dienste sich auf die Loyalität auswirken, verändert sich auch der Kundenwert.

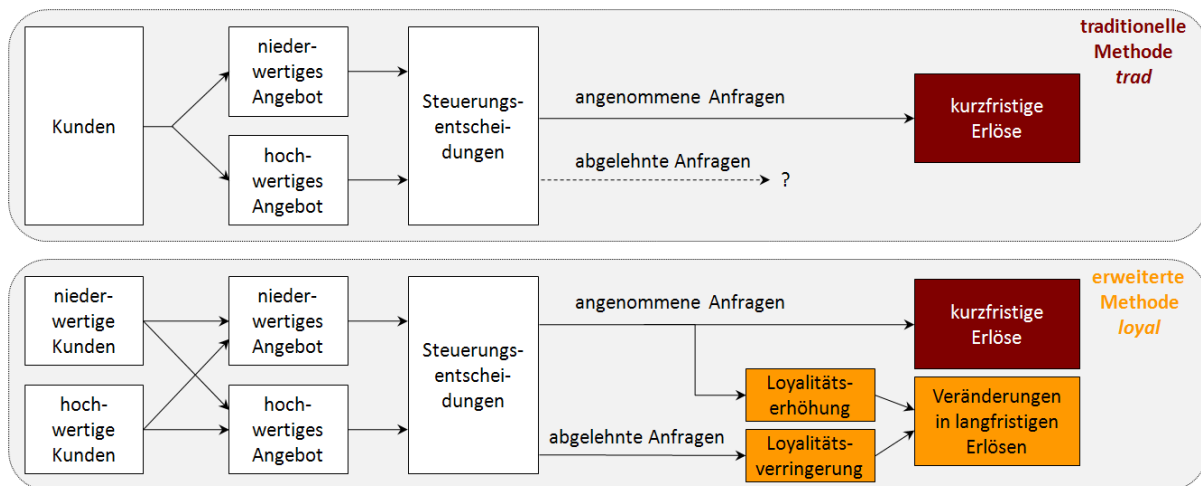


**Bild 1:** Wert des Cloud-Kunden für den Anbieter (in Anlehnung an [6])

Die Verfügbarkeit von IT-Diensten gilt als ein Qualitätskriterium in der Beurteilung des Cloud-Anbieters [8]. Insbesondere im Cloud Computing wird eine nahezu ständige Verfügbarkeit der IT-Dienste durch den Kunden erwartet, da sich diese häufig als kritisch für den Kerngeschäftsbetrieb und den Erfolg der eigenen Dienstleistungen erweisen. Typischerweise werden dafür Service-Level-Agreements zwischen Anbieter und Kunden vereinbart [16]. Da sich der Nutzen für den Cloud-Kunden u.a. aus der dynamischen Skalierung der IT-Ressourcen ergibt [4], ist somit auch die Annahme der angefragten IT-Dienste durch den Anbieter zur Deckung eines Bedarfsanstiegs beim Kunden von enormer Bedeutung. Sowohl der Ausverkauf der Kapazität als auch die Ablehnung der Buchungsanfrage im Rahmen der Erlösmaximierung resultieren für den Kunden in faktischer Verwehrung der Leistungsanspruchnahme, wenngleich letzterer Fall bei gegenteiliger Entscheidung (d.h. Annahme der Anfrage) durch den Anbieter vermeidbar wäre. Studien im Cloud Computing betonen die Wichtigkeit der Verfügbarkeit der (angefragten) IT-Ressourcen zur Sicherstellung der Kundenzufriedenheit und zum Aufbau von Vertrauen zum Anbieter [8], [16]. Andererseits kann die Nichtverfügbarkeit der IT-Dienste in negativen Kundenreaktionen wie Unzufriedenheit, Produkt- bzw. Kategoriewechsel (ggf. anderer Wertigkeit) und Kundenabwanderung münden [3], [4]. Dies ist insbesondere im Cloud Computing (mit Geschäftsbeziehungen im elektronischen Marktumfeld) als kritisch zu bewerten, da hier aufgrund hoher Akquisitionskosten nur treue Kunden langfristig profitabel sind [14]. Letztlich ist daher davon auszugehen, dass die Steuerungsentscheidungen des Anbieters im Rahmen des Revenue Managements (d.h. sowohl Annahme als auch Ablehnung) zu Loyalitätsveränderungen der Kunden führen und damit den Kundenwert erheblich beeinflussen werden (vgl. Bild 1).

Allerdings werden diese erlösrelevanten Loyalitätsveränderungen durch die traditionelle Steuerungsmethode (kurz *trad*) nicht berücksichtigt, da hier nur transaktionsorientierte Erlöse *angenommener* Buchungsanfragen betrachtet werden (vgl. Bild 2, S. 5). Eine differenzierte Zuordnung von Angebot und Kundensegment (von unterschiedlicher Wertigkeit) erfolgt nicht [17]. Im Gegensatz dazu betrachtet die erweiterte Steuerungsmethode (kurz *loyal*) auch Veränderungen in den (langfristigen) Kundenwerten *aller* Buchungsanfragen (inklusive abgewiesenen Kunden), die durch Loyalitätsveränderungen in Folge der Kapazitätsverfügbarkeitsentscheidungen des Anbieters hervorgerufen worden sind. Während die traditionelle Methode *trad* nur kurzfristige Erlöse fokussiert, zieht die erweiterte Methode *loyal* auch zusätzliche Veränderungen in langfristigen Erlösen in Betracht. Die Entscheidung, Kunden mit kurzfristig höherer Zahlungsbereitschaft (aber nicht notwendigerweise die loyalsten Kunden [18]) anzunehmen, sollte Opportunitätskosten in Bezug auf verlorene Kundenwerte einbeziehen, falls im Gegenzug die abgewiesenen

(Potenzial)-Kunden daraufhin ihre zukünftigen Transaktionen reduzieren oder sogar die Beziehung mit dem Anbieter beenden. Das erweiterte Optimierungsmodell unter Berücksichtigung beider Erlösgrößen (kurz- und langfristig, d.h. multikriteriell) soll im nächsten Abschnitt entwickelt werden.



**Bild 2:** Vergleich zwischen traditioneller und erweiterter Steuerungsmethode

## 4 Optimierung und Steuerungsansatz

Das Revenue-Management-System nimmt im Rahmen der Optimierung eine Allokation der (noch) verfügbaren Kapazitätseinheiten auf die erwartete Restnachfrage vor [15]. Die resultierenden Kontingente oder Bid-Preise (als Opportunitätskosten der Kapazitätsnutzung) werden dann der Transaktionssteuerung zugeführt, um über die Annahme bzw. Abweisung der Anfrage zu entscheiden. Diese Entscheidung kann im Hinblick auf die verwendete Steuerungsmethode (*trad* vs. *loyal*) im Revenue-Management-System variieren. Das zu entwickelnde Optimierungsmodell soll hier als Verknüpfung der (formalen) Darstellung des Optimierungsproblems und eines geeigneten Lösungsverfahrens verstanden werden [5].

Es werde angenommen, dass der Cloud-Anbieter über  $h$  Ressourcen verfügt und  $i$  Dienste anbietet. Die Ressourcen (CPU, Arbeitsspeicher, Speicherplatz) sind diskret und werden zu verschiedenen Diensttypen (niedrige, mittlere, hohe Instanzen) gebündelt. Der Bedarf der angebotenen Dienste an diesen Ressourcen entspricht den Amazon Cloud-Parametern und ist in Matrix  $A$  hinterlegt (vgl. Tabelle 1, S. 8; [1]). Das Element  $a_{hi}$  repräsentiert damit die Inanspruchnahme der Ressource  $h$  durch eine Einheit des Dienstes  $i$ , dessen Verkauf einen Erlös  $p_i$  erzielt. Für jede Dienstinstanz wird ein Preisnachlass gewährt, der allerdings mit Restriktionen bezüglich des Betriebssystems einhergeht (vgl. Windows vs. Unix in Tabelle 1, S. 8). Eine Differenzierung bezüglich Bandbreitenbindung (10 vs. 100 Mbit/s), Dienstaktivierungszeit (Minuten vs. Stunden) oder Architektur (32 vs. 64-bit) sind auch denkbar. Die Gesamtkapazität jeder Ressource wird mit  $c_h$  festgelegt. In einem loyalitätsbasiertem Revenue-Management-System müssen die IT-Ressourcen auf die Kombinationen von  $i$  Buchungsklassen (d.h. Diensten) und  $s$  Kundensegmenten (jedes mit bestimmtem Nutzenpotential  $n_s$ ) verteilt werden. Damit können sowohl der Anstieg des Kundenwerts (aufgrund der Loyalitätserhöhung  $I_s^+$  als Folge der Annahme der Anfrage) als auch der Kundenwertverlust (aufgrund der Loyalitätsverringerung  $I_s^-$  als Folge der Ablehnung der Anfrage) modelliert und explizit in der Optimierung berücksichtigt werden.

Die Allokation der Kapazität kann nun als lineares Optimierungsmodell formuliert werden.  $x_{si}^A$  repräsentiert das der Kombination aus Kundensegment  $s$  und Buchungsklasse  $i$  zugewiesene Kontingent [17],  $b_{si}$  (als Element von  $B$ ) stellt die Anzahl bereits vorgenommener Buchungen für eine bestimmte Kombination aus Kundensegment  $s$  und Dienst  $i$  dar,  $y_{si}^A$  repräsentiert die Anzahl abzulehnender Anfragen (von Kundensegment  $s$  für Dienst  $i$ ) und  $d_{sit}$  (als Element von  $D_t$ ) bezeichnet die zu erwartende Restnachfrage in der in Buchungsintervalle  $t \in \{T, \dots, 0\}$  aufgeteilten Buchungsperiode. Der Zielfunktionswert  $Z$  ergibt sich zu:

$$Z(B, D_t) = \max \sum_{s=1}^S \sum_{i=1}^I \left[ (p_i \cdot x_{si}^A) + (l_s^+ \cdot n_s \cdot x_{si}^A) - (l_s^- \cdot n_s \cdot y_{si}^A) \right] \quad (1)$$

$$\text{s.t.} \sum_{s=1}^S \sum_{i=1}^I a_{hi} \cdot (b_{si} + x_{si}^A) \leq c_h \quad \forall h \in \{1, \dots, H\} \quad (2)$$

$$0 \leq x_{si}^A \leq d_{sit} \quad \text{und} \quad 0 \leq y_{si}^A \leq d_{sit} \quad \forall s \in \{1, \dots, S\}; \forall i \in \{1, \dots, I\} \quad (3)$$

$$x_{si}^A + y_{si}^A = d_{sit} \quad \forall s \in \{1, \dots, S\}; \forall i \in \{1, \dots, I\} \quad (4)$$

Die Zielfunktion (vgl. Gleichung (1)) maximiert (über alle Kundensegmente und Buchungsklassen) die Summe aus den gesamten kurzfristigen Erlösen jeder akzeptierten Anfrage plus den daraus resultierenden Gesamtanstieg langfristiger Kundenwerte ( $l_s^+ \cdot n_s$ ) minus die Gesamtverringerung der Kundenwerte ( $l_s^- \cdot n_s$ ) aller abgewiesenen Anfragen als Folge der Steuerungsentscheidungen oder unzureichender Kapazität. Somit sind die möglichen Änderungen der Loyalität für *alle* Buchungsanfragen explizit im Maximierungsproblem integriert. Bei der traditionellen Steuerungsmethode sind  $l_s^+$  und  $l_s^-$  gleich null. Die Zielfunktion  $Z$  unterscheidet sich demnach von den Formulierungen, die stattdessen Loyalitätsmultiplikatoren [17], [18] einbinden. Diese korrigieren den kurzfristigen Erlös um den Kundenwert, um die bevorzugte Verfügbarkeit der Kapazitätsressourcen für loyale Stammkunden abzusichern. Eine Betrachtung der Loyalitätsveränderungen der angenommenen bzw. abgelehnten Kunden kann auf diesem Wege nicht erfolgen. Gleichung (2) stellt sicher, dass die eingesetzte Kapazität für bereits vorgenommene Buchungen und die Kontingente nicht die Gesamtkapazität überschreitet. Es wird des Weiteren festgelegt, dass die Anzahl der zu akzeptierenden als auch abzulehnenden Anfragen nicht-negativ sind und die erwartete Restnachfrage nicht überschreiten (vgl. Gleichung (3)). Auch muss deren Summe der erwarteten Nachfrage im Buchungsintervall entsprechen (vgl. Gleichung (4)).

Die Entscheidung, eine Anfrage anzunehmen oder abzulehnen, basiert auf sogenannten Bid-Preisen. Diese Opportunitätskosten verstehen sich als die Verminderung des erwarteten Gesamterlöses, der mit der Vergabe einer Kapazitätseinheit verbunden ist. Trifft eine Anfrage von Kundensegment  $s$  für Buchungsklasse  $i$  ein (hier durch Matrix  $K$  repräsentiert), wird jeweils die optimale Allokation der noch verfügbaren Kapazität auf die erwartete Restnachfrage berechnet, wobei dieses Optimierungsproblem einmal *mit* und einmal *ohne* Berücksichtigung der aktuellen Anfrage gelöst wird [17]. Die Annahme der Anfrage vermindert dabei die noch verfügbare Restkapazität. Der resultierende Zielfunktionswert  $Z(B+K, D_{t-1})$  als Restwert der Kapazität kann sich vom Fall der Ablehnung ( $Z(B, D_{t-1})$ ) unterscheiden. Die Anfrage wird im Rahmen der Transaktionssteuerung im Revenue-Management-System immer dann angenommen, solange die Summe aus transaktionsorientiertem Erlös und möglichen Änderungen im langfristigen Wert des Kunden (sowohl Anstieg als auch Verringerung) die Opportunitätskosten mindestens ausgleicht (vgl. Gleichung (5)). D.h. der Zielfunktionswert  $Z(B+K, D_{t-1})$  im Falle der Annahme der Anfrage plus die Summe aus transaktionsorientiertem Erlös und Kundenwertanstieg muss zumindest

genauso groß sein wie  $Z(B, D_{t-1})$  im Ablehnungsfall minus den resultierenden Kundenwertverlust (vgl. Gleichung (6)).

$$p_i + (l_s^+ \cdot n_s) + (l_s^- \cdot n_s) \geq Z(B, D_{t-1}) - Z(B + K, D_{t-1}) \quad (5)$$

$$\Leftrightarrow Z(B + K, D_{t-1}) + p_i + (l_s^+ \cdot n_s) \geq Z(B, D_{t-1}) - (l_s^- \cdot n_s) \quad (6)$$

## 5 Simulation und Evaluation

Im Rahmen dieses Abschnitts soll das loyalitätsbasierte Revenue-Management-System evaluiert werden. Hierfür ist der Einsatz von Simulationen verbreitet [2], da sie eine große Zahl an Variablen berücksichtigen können und es erlauben, bestimmte Entscheidungssituationen unter Maßgabe von streuungs- bzw. zufallsbehafteten Daten mit verschiedenen Parameterwerten durchzurechnen und die Auswirkungen auf das Ergebnis aufzuzeigen [10]. Simulationen bieten damit trotz hoher Rechenintensität auch im Umfeld des Revenue Managements eine aussichtsreiche Möglichkeit, Zusammenhänge in komplexen Nachfrageumgebungen zu analysieren und die Steuerungsentscheidungen dahingehend auszurichten.

### 5.1 Simulationsziel, -umgebung und -design

Grundsätzliches Ziel der Simulation liegt im Vergleich der beiden Kapazitätssteuerungsmethoden *trad* vs. *loyal*. Die erzielten Erlöse werden dabei sowohl untereinander als auch mit der trivialen First-Come-First-Serve-Steuerung (kurz *fcfs*) verglichen. Im modellierten Anwendungsszenario im Umfeld des Cloud Computing werden die vom Anbieter angebotenen IT-Dienste von den Cloud-Kunden vornehmlich zum Hosting ihrer eigenen Webseiten und Onlineshops genutzt. Die Kapazitätssteuerung soll in einem Zeitraum erfolgen, der durch eine Übernachfrage (im Vergleich zur verfügbaren Kapazität des Anbieters) nach IT-Diensten (z.B. einige Wochen vor Weihnachten) gekennzeichnet ist. In diesem endlichen Zeithorizont  $T$  ist durch die gestiegenen Zugriffszahlen durch die Kunden der Cloud-Kunden ein gesteigerter Bedarf nach IT-Diensten beim Cloud-Anbieter zu erwarten, der danach erfahrungsgemäß wieder abflacht. Nutzt der Cloud-Anbieter die IT-Dienste auch selbst (wie z.B. Amazon), kann es in praxi tatsächlich zu einer Lastspitze kommen und nicht alle Anfragen nach weiteren IT-Diensten können gewährt werden.

Die beschriebene Bid-Preis-Steuerung wurde in MATLAB 7.0 von Mathworks Inc. implementiert. Um durch die Simulation eine angemessene Aussagefähigkeit zu erlangen [10], sollen vor allem der im Revenue Management als relevant erachtete Netzwerkcharakter (d.h. Bündelung mehrerer Ressourcen zu einem Angebot), eine entsprechend umfassende Zahl an Buchungsklassen und eine stochastische Nachfrageverteilung Berücksichtigung finden [17]. Während der Nachfrageumfang nach den verschiedenen IT-Diensten deterministisch ist, erfolgen die Ankunft der Buchungsanfragen, die Zuordnung der angebotsbezogenen Anfragen zu Kundensegmenten sowie die tatsächliche Ausprägung des Nutzenpotentials und der Loyalitätsveränderungen stochastisch. Die Nachfrage für jeden der sechs angebotenen IT-Dienste ist unabhängig voneinander, d.h. ein Kunde, der ein bestimmtes Angebot anfragt, wird kein anderes buchen. Die Ankunft der Anfragen wird durch einen homogenen Poisson-Prozess mit konstanten Ankunfts-raten für jedes Angebot modelliert. Es wird angenommen, dass Anfragen für niederwertige Angebote (d.h. niedriger Preis bzw. Unix) eher früher und Anfragen nach höherwertigen Angeboten (d.h. hoher Preis bzw. Windows) eher später eintreffen [2]. Insgesamt wurden 5000



verschiedene Ankunftsprozesse simuliert, die in Nachfrageschwankungen resultieren. D.h. die tatsächliche Nachfrage kann aufgrund des zufälligen Eintreffens der Anfragen von der erwarteten Nachfrage abweichen (vgl. Streuung der Nachfrage (Standardabweichung) über alle Durchläufe in Tabelle 1). Da jedoch keine Prognosefehler modelliert werden, entspricht die in der Zielfunktion  $Z$  verwendete, prognostizierte Nachfrage in jedem Simulationsdurchlauf der realisierten Nachfrage.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$c_h$
CPU	1	1	4	4	8	8	600
Arbeitsspeicher (GB)	2	2	7	7	15	15	1000
Speicherplatz (GB)	160	160	850	850	1690	1690	110000
Betriebssystem	Windows	Unix	Windows	Unix	Windows	Unix	
Preis	12	10	48	40	96	80	
erwartete Nachfrage	60	60	45	45	30	30	
Streuung Nachfrage	7,7	7,8	6,6	6,6	5,5	5,5	

**Tabelle 1: Ressourcen-, Preis- und Mengenmodell**

Grundsätzlich lassen sich im Simulationsdesign zwei Parameterformen (*real* vs. *angenommen*) unterscheiden. Im (simulierten) Revenue-Management-System wird daher nur eine Annahme über die realen Parameterausprägungen getroffen und die Kapazitätssteuerung anhand der angenommenen Parameterwerte vorangetrieben. Die Unterscheidung der Parameterformen betrifft im Rahmen der Simulation das Nutzenpotential und die Veränderungen der Kundenloyalität aufgrund der Steuerungsentscheidungen. Dies soll dem Umstand Rechnung tragen, dass durch den Anbieter nur Annahmen über die tatsächlichen Ausprägungen jener unsicherheitsbehafteten Größen angestellt werden können. Aus Gründen der Berechnungskomplexität werden diese kundenindividuellen Größen im Optimierungsmodell des Revenue-Management-Systems auf Kundengruppenebene zusammengefasst. Um verschiedenen langfristigen Wertbeiträgen der Kunden zu genügen, wurden ein hoch- und niederwertiges Kundensegment (d.h. angenommenes Nutzenpotenzial  $n_1 = 400$ ,  $n_2 = 200$ ) modelliert. Das tatsächliche Nutzenpotenzial des anfragenden Kunden wird durch eine gleichverteilte Zufallsziehung im entsprechenden Intervall ermittelt, d.h.  $\hat{n}_1 \in [300; 500]$  und  $\hat{n}_2 \in [100; 300]$ . Die durchschnittliche Beziehung zwischen Zahlungsbereitschaft (d.h. angefragtem IT-Dienst) und langfristigem Wert des Kunden (d.h. anfragendes Kundensegment) wird durch  $r \in [-1; 1]$  abgebildet. Ein stark positiver Zusammenhang bedeutet, dass höherwertige (niederwertige) Kunden stets die höherwertigen (niederwertigen) Angebote nachfragen, bei einer stark negativen Relation werden stattdessen die höherwertigen (niederwertigen) Angebote immer von niederwertigen (höherwertigen) Kunden nachgefragt [17]. Beim gewählten  $r = 0$  ergibt sich die erwartete Nachfrage nach einem Angebot damit zu gleichen Teilen aus beiden Kundensegmenten. Die Werte der Loyalitätsveränderungen werden mit  $l_1^+ = l_2^+ = 0,01$  und  $l_1^- = l_2^- = 0,05$  angenommen. Eine Ablehnung der Anfrage führt damit zu einer größeren Loyalitätsveränderung als im Fall der Annahme. Die reale Veränderung der Loyalität des jeweiligen Kunden wird abermals per gleichverteilter Zufallszahl entschieden, d.h.  $\hat{l}_1^+, \hat{l}_2^+ \in [0; 0,02]$  und  $\hat{l}_1^-, \hat{l}_2^- \in [0,04; 0,06]$ .

In der Praxis kann zur Schätzung der Parameter zunächst die Analyse der vorangegangenen Transaktionen als Ausgangspunkt dienen. Da im Cloud Computing ein hoher Geschäftskundenanteil auszumachen ist [8], können die anfragenden Kunden oftmals durch Kunden-

nummer bzw. Login eindeutig identifiziert und so die kundenindividuelle Historie aus der Datenbank des Revenue-Management-Systems (CRM-Schnittstelle) abgerufen werden, die Aussagen zur Kundenwertigkeit (z.B. Anhaltspunkte zur Prognose zukünftiger Bestellungen) erlaubt. Durch Methoden des Data Mining können zudem Kundencluster mit ähnlichen Merkmalen gebildet werden [8]. Insbesondere die Analyse der Entwicklung des Kaufverhaltens von Kunden mit vergleichbaren bisherigen Ausgaben aber unterschiedlichen Steuerungsentscheidungen bezüglich ihrer Anfragen kann aufschlussreiche Erkenntnisse liefern. Angereichert mit Ergebnissen von durchgeführten Befragungen [16] zu geplanten Ausgaben der Kunden, zum Anteil der Ausgaben beim Anbieter an den Gesamtausgaben [6] und zu möglichen Kundenreaktionen auf Steuerungsentscheidungen können so Rückschlüsse auf das zu schätzende Nutzenpotential und Loyalitätsänderungen ermöglicht werden.

## 5.2 Simulationsergebnisse

Zur Bewertung der verschiedenen Steuerungsmethoden wurden die Ergebnisgrößen Gesamttransaktionserlöse  $e^{trans}$ , Gesamtkundenwerterhöhung bzw. -verringerung  $e^{cv+}$  bzw.  $e^{cv-}$  und Gesamterlöse  $e^{gesamt} = e^{trans} + e^{cv+} - e^{cv-}$  (gemittelt über alle Simulationsdurchläufe) definiert (vgl. Tabelle 2). Die Steuerungsentscheidungen erfolgen im Rahmen des Opportunitätskostenvergleichs auf Basis der angenommenen Erlöse (bedingt durch die angenommenen Werte für  $n_s$ ,  $I_s^+$  und  $I_s^-$ ), der Vergleich der Methoden bezieht sich allerdings auf die tatsächlich generierten Erlöse (bedingt durch die kundenindividuell verschiedenen Werte von  $\hat{n}_s$ ,  $\hat{I}_s^+$  und  $\hat{I}_s^-$ ). Diese Wertschwankungen erklären auch die insgesamt hohe Streuung (Standardabweichung) bei  $e^{gesamt}$ .

Kennzahl	<i>fcfs</i>	<i>trad</i>	<i>loyal</i>	<i>loyal / trad</i>
$e^{gesamt}$	4744	5349	5994	1,121
$e^{trans}$	5661	6234	6116	0,981
$e^{cv+}$	521	526	654	1,243
$e^{cv-}$	1438	1411	776	0,55
Streuung $e^{gesamt}$	341	305	144	0,472

**Tabelle 2: Simulationsergebnisse**

Auch wenn die First-Come-First-Serve-Steuerung ähnlich viele Anfragen wie *trad* annimmt, weist sie aufgrund der fast ausnahmslosen Annahme von niederwertigen Angebotsanfragen ein schwaches Ergebnis in  $e^{trans}$  auf, da diese Anfragen am Beginn der Buchungsperiode eintreffen. Sehr hoch (aber nur marginal schlechter verglichen mit *trad*) fällt  $e^{cv-}$  aus, weil alle nachfolgenden Anfragen wegen der bereits erschöpften Kapazität abgelehnt werden müssen – unabhängig vom anfragenden Kundensegment und der Auswirkung auf die Loyalität.

Die Simulationsergebnisse bestätigen die Annahme, dass loyalitätsbasiertes Revenue Management im Vergleich zur traditionellen Steuerung zwar in niedrigeren kurzfristigen Erlösen ( $e^{trans}$ ) aber dafür höheren langfristigen Erlösen mündet. Insgesamt erzielt *loyal* durchschnittlich 12,1% höhere Gesamterlöse als *trad* (vgl. Tabelle 2). Es ist anzumerken, dass *loyal* mehr Kundenanfragen (von beiden Kundensegmenten) annimmt als *trad*. Dies führt zu einem höheren  $e^{cv+}$  und niedrigeren  $e^{cv-}$ . Im Gegensatz zur traditionellen Kapazitätssteuerung lehnt *loyal* viele Anfragen von niederwertigen Kunden für das preis- und ressourcenintensive Angebot  $i = 5$  ab. Damit

toleriert *loyal* den Verlust in höheren Transaktionserlösen zugunsten freier Kapazität für die Vergabe an viele Nachfrager (aus beiden Kundensegmenten) für den IT-Dienst  $i = 2$  und an höherwertige Kunden für  $i = 4$ . Diese zusätzlich akzeptierten Anfragen werden einen positiven Einfluss auf die erwarteten Kundenwerte in zukünftigen Buchungsperioden haben.

### 5.3 Simulationsergebnisse bei Kundenabwanderung

Die Entscheidung zur Angebotsverfügbarkeit hat vor allem dann eine hohe Tragweite, wenn durch die Ablehnung der Anfrage eine Abwanderung des Kunden droht und damit zum Verlust des Kundenwerts für den Anbieter führt. Diese kann in Praxis dann erwartet werden, wenn der Kunde bereits mehrfach (hintereinander) abgelehnt wurde. Die erneute Verwehrung der benötigten IT-Ressourcen kann als Folge die Erstellung der eigenen Leistungen des Cloud-Kunden gefährden (siehe beschriebenes Anwendungsszenario) und tatsächlich in einer Kundenabwanderung münden. Die Berücksichtigung dieser negativen Effekte soll nun im Optimierungsmodell und in der Transaktionssteuerung vollzogen werden.

Der Anteil der Anfragen jener abwanderungsgefährdeten Kunden zu allen eintreffenden Anfragen soll  $f_s = 0,05$  betragen, wobei der Anbieter diese unsicherheitsbedingt mit einer Erkennungsrate von  $j_s = 0,5$  auch als solche einstuft. Im Fall der Ablehnung wird angenommen, dass sich der Loyalitätsverlust von  $l_s^-$  auf  $g_s^- = 0,25$  erhöht. Der Parameter  $g_s^-$  kann damit als Schwellwert interpretiert werden, bei dem die aktuelle, durch die erneute Ablehnung verringerte Loyalität soweit abgesunken ist, dass der Kunde die Beziehung mit dem Anbieter beendet. Der reale Loyalitätsverlust bestimmt sich durch gleichverteilte Zufallsziehung, d.h.  $\hat{g}_s^- \in [0,2; 0,3]$ . Zur entsprechenden Berücksichtigung im Optimierungsmodell wird daher jedes Kundensegment (hoch- vs. niederwertig) in zwei Untersegmente aufgeteilt: Kunden mit (bzw. ohne) Gefahr einer Kundenabwanderung infolge der Ablehnung und somit hohem (bzw. nur geringem) Kundenwertverlust. Dies ist erforderlich, um die unterschiedlichen Werte der (anteiligen) Restnachfrage (d.h.  $f_s$  vs.  $1 - f_s$ ) und Loyalitätsverringeringen im Ablehnungsfall (d.h.  $l_s^- = 0,05$  vs.  $l_s^- = g_s^- = 0,25$ ) abzubilden. Der Restwert der Kapazität  $Z$  wird daher im Optimierungsmodell nun auch durch abwanderungsgefährdete Kunden beeinflusst (vgl. Gleichung (1)). Erkennt der Anbieter die drohende Kundenabwanderung (gesteuert über  $j_s$ ), wird die erweiterte Steuerungsmethode *loyal* im Rahmen der Transaktionssteuerung die Loyalitätsverringering mit  $l_s^- = g_s^-$  beziffern (vgl. Gleichung (5)). Dies soll die bevorzugte Verfügbarkeit der Kapazitätsressourcen für den Kunden absichern, um den drohenden Kundenwerttotalverlust zu vermeiden.

Die Simulationsergebnisse können mit denen aus Abschnitt 5.2 verglichen werden, da die Steuerungsmethoden *fcfs*, *trad* und *loyal* in jedem simulierten Szenario dem identischen Anfragestrom unterliegen (vgl. relative Erlösveränderung zur Simulation ohne Kundenabwanderung als Klammerwerte in Tabelle 3, S. 11). Aufgrund unverändertem Optimierungsmodell und Transaktionssteuerung unterscheiden sich die jeweiligen Steuerungsentscheidungen bei *fcfs* und *trad* nicht, d.h.  $e^{trans}$  und  $e^{cv+}$  bleiben gleich. Die Gesamterlöse sinken stattdessen bei allen Methoden wegen einem erhöhten  $e^{cv-}$  infolge der eingeführten Kundenabwanderung. Auch wenn *loyal* die Abwanderung nicht zuverlässig erkennen kann, verringert sich  $e^{gesamt}$  nur um ca. 1,7% (hingegen 5,2% bei *trad*) im Vergleich zur Simulation in Abschnitt 5.2. Der geringere Verlust ist durch den wesentlich niedrigeren Anstieg in  $e^{cv-}$  (11,5% vs. 19,8% bei *trad*) bedingt. Die marginal geringeren Transaktionserlöse bei *loyal* sind durch die Toleranz ggü. Kombinationen aus

nachgefragtem Angebot und Kundensegment, die nur aufgrund der drohenden Kundenabwanderung anstelle lohnenderer Kombinationen angenommen werden, zu erklären. Die Ergebnisse in Tabelle 2, S. 9 und Tabelle 3 zeigen zudem, dass die Gesamterlöse von *loyal* einer wesentlich geringeren Streuung (Standardabweichung) über alle Simulationsdurchläufe verglichen mit *trad* unterliegen. Damit zeichnet sich die erweiterte Methode neben generell höheren langfristigen Erlösen auch durch eine Robustheit ggü. den zufallsbedingten, für das Revenue-Management-System in ihrer realen Ausprägung unbekannten Werten ( $\hat{r}_s, \hat{I}_s^+, \hat{I}_s^-, \hat{g}_s^-$ ) aus.

Erlöse	<i>fcfs</i>	<i>trad</i>	<i>loyal</i>	<i>loyal / trad</i>
$e^{gesamt}$	4455 (-6,1%)	5070 (-5,2%)	5895 (-1,7%)	1,163
$e^{trans}$	5661 (0%)	6234 (0%)	6107 (-0,1%)	0,98
$e^{cv+}$	521 (0%)	526 (0%)	652 (-0,3%)	1,24
$e^{cv-}$	1727 (+20,1%)	1690 (+19,8%)	865 (+11,5%)	0,512
Streuung $e^{gesamt}$	418	390	178	0,456

**Tabelle 3: Simulationsergebnisse mit Kundenabwanderung**

Die Ergebnisse der Simulationen lassen darauf schließen, dass loyalitätsbasiertes Revenue Management dann angebracht ist, wenn durch die Steuerungsentscheidungen des Anbieters (Annahme und Ablehnung) eine Auswirkung auf die Kundenloyalität zu erwarten ist. Dafür ist eine Klassifikation der Anfragen in  $s$  Kundensegmente und angemessene Annahmen über  $I_s^+$  und  $I_s^-$  notwendig. Die Anwendung des erweiterten Revenue-Management-Systems ist sinnvoll in Zeiten hoher Nachfrage und falls hochwertige Anfragen spät eintreffen.

## 6 Zusammenfassung und Ausblick

Durch die Berücksichtigung der Auswirkungen der Steuerungsentscheidungen ermöglicht das loyalitätsbasierte Revenue-Management-System sowohl die effiziente Kapazitätsnutzung als auch den Aufbau langfristig profitabler Kundenbeziehungen. Die Simulationsergebnisse unterstreichen die Vorteile der erweiterten Optimierung und Transaktionssteuerung. Forschungsbedarf besteht in der Schätzung der Loyalitätsveränderungen und der Identifizierung von geeigneten Modellen zur Vorhersage des Nutzenpotenzials, um eine Kundenbewertung vollziehen zu können. Es erscheinen zudem weitere Simulationen mit geänderten Parametern (Ressourcen-, Preis-, Mengenmodell mit Prognosefehlern) sinnvoll. So kann auch eine Sensitivitätsanalyse mit veränderten Werten von  $g_s^-, I_s^+, I_s^-$  (für jedes Kundensegment) neue Erkenntnisse zur Stabilität der Ergebnisse [10] liefern. Die Integration von Kapazitätsteuerung und Kundenbeziehungsmanagement bleibt ein vielversprechendes Forschungsgebiet mit signifikanten Implikationen zur Verbesserung der Wettbewerbsfähigkeit (nicht nur) von Cloud-Anbietern.

## 7 Literatur

- [1] Amazon.com (2011): Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/de/ec2/> Abgerufen am 21.09.2011.
- [2] Anandasivam, A; Buschek, S; Buyya, R (2009): A Heuristic Approach for Capacity Control in Clouds. In: *Proceedings of the 11th IEEE Conference on Commerce & Enterprise Computing*, 90-97.
- [3] Anderson, ET; Fitzsimons, GJ; Simester, D (2006): Measuring and Mitigating the Costs of Stockouts. In: *Management Science* 52(11):1751-1763.
- [4] Armbrust, M; Fox, A; Griffith, R; Joseph, AD; Katz, RH; Konwinski, A; Lee, G; Patterson, DA; Rabkin, A; Stoica, I; Zaharia, M (2009): Above the Clouds - A Berkeley View of Cloud Computing. UC Berkeley Technical Report UCB/EECS-2009-28.
- [5] Fink, A; Schneidereit, G; Voß, S (2005): Grundlagen der Wirtschaftsinformatik. 2. Auflage, Physica-Verlag, Heidelberg.
- [6] Gupta, S; Hanssens, D; Hardie, B; Kahn, W; Kumar, V; Lin, N; Ravishanker, N; Sriram, S (2006): Modeling Customer Lifetime Value. *Journal of Service Research* 9(2):139-155.
- [7] Hippner, H (2006): CRM - Grundlagen, Ziele und Konzepte. In: Hippner, H; Wilde, KD (Hrsg.): *Grundlagen des CRM*. 2. Auflage, Gabler, Wiesbaden, 15-44.
- [8] Koehler, P; Anandasivam, A; Dan, MA; Weinhardt, C (2010): Customer Heterogeneity and Tariff Biases in Cloud Computing. In: *Proceedings of ICIS 2010*, Paper 106.
- [9] Lassmann, W (2006): Wirtschaftsinformatik. Gabler, Wiesbaden.
- [10] Lehneis, A (1970): Die Reduktion der Unsicherheit der Erwartungen im Rahmen der langfristigen Planung der Unternehmung. Diss., Universität Erlangen-Nuernberg.
- [11] Lindenmeier, J (2005): Yield-Management und Kundenzufriedenheit. Diss., Universität Freiburg.
- [12] Mechler, B (1995): Intelligente Informationssysteme, Addison-Wesley, Bonn.
- [13] Pueschel, T; Anandasivam, A; Buschek, S; Neumann, D (2009): Making Money with Clouds - Revenue Optimization through automated Policy Decisions. In: *Proceedings of the 17th European Conference on Information Systems*, 355-367.
- [14] Reichheld, FF; Scheffer, P (2000): E-loyalty - Your Secret Weapon on the Web. In: *Harvard Business Review*, Jul-Aug:105-113.
- [15] Talluri, KT; van Ryzin, GJ (2004): The Theory and Practice of Revenue Management. Springer, Berlin.
- [16] Uusitalo, I; Karppinen, K; Juhola, A; Savola, R (2010): Trust and Cloud Services - An Interview Study. In: *Proceedings of CLOUDCOM 2010*, 712-720.
- [17] von Martens, T (2009): Kundenwertorientiertes Revenue Management im Dienstleistungsbereich. Gabler, Wiesbaden.
- [18] Wirtz, J; Kimes, SE; Pheng Theng, JH; Patterson, P (2003): Revenue Management - Resolving Potential Customer Conflicts. In: *Journal of Revenue & Pricing Management* 2(3):216-226.